Date: December 16, 2010

To: B. Conetta (USEPA Region 2)

From: E. Garvey, J. Atmadja, C. Prabhu, X. Wang (The Louis Berger Group, Inc.), J. Kern (Kern Statistical Services, Inc.), and D. Hayes (University of Louisiana at Lafayette)

Subject: Issues in Estimation of Depth of Contamination and Application to Dredge Prism

# 1. THE ROLE OF UNCERTAINTY IN ESTIMATING DEPTH OF CONTAMINATION

## 1.1. Overview

**Issue:** Based on performance of Phase 1dredging and Peer Reviewers' recommendations, GE has proposed adding 6 inches to their Phase 2 design dredge depths in thin PCB deposits with estimated depth of contamination (DoC) less than 18 inches and adding 4 inches to thicker deposits with estimated DoC exceeding 18 inches.

**EPA Position:** It is EPA's position that GE's approach is unlikely to adequately improve characterization of the elevation necessary to identify the depth of all PCB deposits with concentrations exceeding 1 mg/kg total PCBs as required by the record of decision (ROD; USEPA, 2002).

**Resolution:** EPA proposes a performance-based approval of GE's dredging design that will incorporate specific metrics that must be attained during each dredging season and at the completion of dredging (see Revised EPS for Phase 2). EPA will retain authority to require that corrective actions be taken during or after a dredge season to ensure that these metrics will be met. EPA has had extensive past and recent discussions with GE in which EPA emphasized the important negative effects of uncertainty in DoC (both measured and modeled) on performance of the dredging design. During development of the Phase 1 design, GE's position was that core segmentation, extrapolation of incomplete cores, interpolation methods, and identification of geologically confining layers (*i.e.*, clay and bedrock) were all conservative with regard to estimation of DoC. In retrospect, the Phase 1 experience showed that collectively these factors were not adequately conservative to compensate for the large uncertainty in DoC, primarily due to uncertainties of extrapolated PCB concentrations in incomplete cores, inaccuracies in the determination of absolute elevation of core samples, and uncertainty of predicted DoC at unsampled locations.

Although GE is collecting additional cores at incomplete core locations, it is EPA's contention that substantial uncertainties are not understood in Phase 2 areas based on data provided by GE to date. Although EPA believes there is a strong likelihood of substantial uncertainty in DoC determination, GE has argued against quantification of expected uncertainties in determination of DoC for Phase 2 dredge prism development. Under these circumstances, EPA has chosen to adopt specific performance requirements utilizing limits on area capped as capping performance criteria with well-defined adaptive management options that can be applied early in Phase 2 - Year 1 (2011). If performance criteria are not met within specific time frames, adaptive management processes will be triggered for subsequent CUs.

## 1.2. Supporting Rationale

Dredging of unexpected PCB inventory in the Phase 1 CUs was necessary because the design cut lines based on GE's deterministic model underestimated the true DoC. Overall, if design volumes are adjusted for the physical offsets adjacent to structures that were necessary to manage sediments at the shoreline, the amount dredged in the 10 Phase 1 CUs was nearly double the originally planned volume. This additional unplanned dredging significantly affected Phase 1 with respect to compliance with all three performance standards (Bridges *et al.*, 2010; pp. iii and v) The relevance and consequences of uncertainty in the DoC measurements was a point of disagreement between EPA and GE over the course of the remedial design. In the comments and associated responses that were exchanged between EPA and GE regarding the Intermediate Design Report (IDR) in December of 2005, EPA warned GE that the uncertainty at individual core locations, which EPA estimated to be about 1 foot, would outweigh GE's contention that DoC measured in the cores would be conservative; EPA contended that "underestimating DoC may lead to additional re-dredging to remove inventory."

GE responded to these concerns in the intermediate design report (GE, 2005) stating that "GE believes that the uncertainty derived from Kriging cannot be used to infer the true uncertainty of the DoC surface so that the bulleted information above can be calculated with confidence." The referenced bulleted information consisted of a set of metrics that EPA recommended for GE to quantify, including re-dredging rates, proportion of inventory left in place, proportion of "clean" sediment to be removed, uncertainty in total volume, percentage of CUs that might fail the performance standard, and potential implications to schedule. These were all items that EPA suggested could be evaluated through adopting statistical approaches to evaluate uncertainty. In lieu of efforts to quantify components of DoC uncertainty that EPA identified, GE asserted, "**It is GE's view that such an attempt would be meaningless and the only reasonable way to obtain this information is through the Phase 1 dredging program**" (GE, 2005, Attachment D, p. 6-2; emphasis added).

In light of GE's consistent adherence to a deterministic approach to DoC development, EPA adopted a performance-based approach in approving the Phase 1 design and allowed GE flexibility to manage the uncertainty in DoC through other means as they implemented the project.

Phase 1 dredging results validate EPA's concerns regarding uncertainty in DoC estimates, as the average thickness of additional dredging required was greater than 1 foot and ranged to more than 9 feet. Underestimation of the DoC resulted in significant re-dredging to remove inventory and not residuals. This also resulted in multiple re-dredging passes which adversely impacted resuspension and project schedule. Since GE clearly considered its design sufficiently robust to deal with DoC uncertainty during implementation, it is not appropriate to consider the consequences of this design flaw as an unexpected impediment to productivity, nor to attribute the need to re-dredge to the Residuals Standard.

For the Phase 2 - Year 1 (2011) dredging season, GE has proposed to use the same modeling methods for DoC with the addition of an overcut of 6 and 4 inches in thin and thick deposits, respectively, arguing that this addition to DoC will mitigate uncertainty in the dredge design elevations. Given that average understatement of DoC varied spatially with an average of more than a foot, and ranged to more than 9 feet, EPA does not believe that this proposition is a reasonable approach to aggressively targeting the 1 mg/kg total PCBs elevation as recommended by the Peer Review Panel. Because EPA and GE strongly disagree on the approach to targeting DoC, EPA has decided to allow GE to implement its approach and to ensure that it can be corrected if unsuccessful; EPA has decided to adopt a performance-based approval of the GE design cut lines, this time with the imposition of additional performance metrics and several evaluation milestones.

EPA has considered four lines of evidence in development of its position. These lines of evidence are based on 1) reluctance to mitigate EPA's concerns regarding the negative effects of uncertainty on DoC estimates, 2) diagnostic testing performed on GE's DoC model, 3) observations of GE's dredge design performance in 2009, and 4) new data collected in 2010 intended to support subsequent Phase 2 designs. Following is a summary of these four lines of evidence and a discussion leading to EPA's position regarding development of dredging design elevations.

### 1.3. Inadequate Compensation for Known Uncertainties

GE inadequately compensated for known uncertainties in DoC characterization identified by both EPA and GE in 2004 and 2005, and did not heed EPA's warnings that the failure to compensate for uncertainty would likely increase the frequency of re-dredging efforts necessary to remove PCB inventory. GE developed a dredging design that did not incorporate these known uncertainties identified by EPA, including 1) DoC extrapolation error, 2) DoC interpolation error, 3) clay layer delineation error, 4) coarse segmentation of cores, and 5) the known dimensions of the previously dredged navigation channel.

<u>DoC Extrapolation</u>

GE has reported that extrapolation of depth of contamination based on incomplete cores is an important source of error in development of the Phase 1 design elevations. Benaman, *et al.* (2010) stated "The PCB profiles in these residual cores showed that the accuracy of the cut line was highly sensitive to the data coverage and to the quality of the design data, specifically whether the full PCB profile was captured in the design sediment core." EPA agrees that DoC extrapolation is error-prone and was an important factor contributing to high Phase 1 re-dredging rates. Errors in extrapolation of PCB concentration was also a source of error contributing to underestimation of the total in-place volume and mass of contaminated sediment and subsequent difficulties meeting the Residual Standard.

Appendix H of GE's Phase 1 Evaluation Report (GE, 2010; Page H-2) states, "During the Phase 1 DAD, the EPA established an extrapolation equation[1] to determine the profile of PCBs below the last measured section for any core whose PCB profile did not decline to less than 1 mg/kg Total PCBs (and therefore the depth to clean sediment had not been determined from measured data)." The footnote (1) included in the quote refers to Kern Statistical Services (2004) entitled "*Statistical Analysis of DOC Extrapolation Data*" from which GE incorrectly claims that EPA "established" an extrapolation equation.

On the contrary, the referenced memo reports the findings of a comparative study <u>evaluating the uncertainty</u> of candidate extrapolation models. The study was conducted in response to a GE proposal to extrapolate incomplete cores using an exponential decay rate of (-0.24 mg/kg-inch). It demonstrated that the method adopted by GE might entail extrapolation errors as large as +/-34 inches and would in the facsimile population developed from Level 1A (*i.e.*, complete) cores.

Kern (2004) also evaluated models that penalized understatement of DoC more strongly than overstatement, but GE elected not to employ such methods. Rather, the model selected by GE to extrapolate DoC from incomplete cores in Phase 1 areas had a rate coefficient of -0.186 mg/kg-inch. This was slightly lower than the -0.24 mg/kg- inch coefficient originally proposed by GE, which resulted in marginally deeper DOC extrapolations, but much shallower extrapolations than would have resulted from the other models EPA evaluated based on stronger penalties for underestimation.

By their selection, GE demonstrated that they valued under-estimation of DoC as being of no greater importance than over-estimation. The consequences of that value judgment by GE led to significant understatement of mass and volume of the PCB inventory (by nearly half, as noted above), large numbers of re-dredging passes, and the inability to meet the residual engineering performance standard in Phase 1. GE also did not incorporate uncertainty in extrapolated values into their interpolation of DoC, by treating extrapolated and actual data values as if they were equally accurate. GE makes no mention of the potential for 34-inch errors in extrapolated values that were identified by Kern (2004) in either the Phase 1 DAD or Appendix H of their Phase 1 Evaluation Report.

In Appendix H, GE did not acknowledge that Kern (2004) provided the following caution and recommendations regarding the applicability of the candidate extrapolation models to actual incomplete cores:

> *"Analyses reported here were based on complete cores where the approximate true depth of contamination was known. The results are to be applied to incomplete cores**. It is not know[n] how well this analysis may apply to the actual population of incomplete cores**. It is my understanding that additional co-located cores are to be collected at the incomplete core locations. These pairs of co-located cores will provide additional data that could be used to re-estimate the parameters of equation (1). I would recommend that before incomplete cores are extrapolated, these new data be used to re-estimate the rate parameter and to test the level of agreement and bias between extrapolated and true depth of contamination**. Development of final analyses should <u>not</u> proceed until these new data are available and the core extrapolation results are fully evaluated.**"* (Emphasis added).

This recommendation points out that the complete cores are unlikely to represent the same physical setting, and subsequently DoC, as the actual incomplete cores to which the extrapolation was to be applied. Because the test population was composed of complete cores, they were preferentially biased toward areas where such cores were more likely to have been obtained given the coring method used. These include thinner deposits, areas with less debris, and areas with layers of lower PCB concentration inter-burden. These preferential biases would be expected to cause the apparent decay rate to be higher than might be necessary to extrapolate actual incomplete cores.

It was clearly stated that extrapolation models should not be applied without validation and re-estimation based on actual pairs of co-located complete and incomplete cores upon their becoming available. If such an analysis was conducted by GE, it was not reported to EPA. GE applied the coefficients derived from the simulation study as though they were applicable to actual incomplete cores, and subsequently treated extrapolated data as though they were sample values. This turned out to be a critical flaw in development of the Phase 1 design. .

DoC Interpolation Error

After publishing the Phase 1 DAD Report that included substantial statistical error analysis, GE developed an alternative deterministic method for interpolation of PCB concentrations based on inverse distance weighting interpolation. The method involved interpolation of sample concentrations, extrapolated estimates, and imputed zero concentration values within each of seventeen, six-inch intervals starting from the sediment surface and working down to the bottom layer at a depth of 8.5 feet. . Note actual dredging extended to as much as 13 feet, a depth that no one anticipated in Phase 1 CUs. These layers were stacked and searched to identify the depth within each interpolated cell (5-x-5 foot grid) at which the DoC was identified.

In the Phase 1 DAD, GE also reported a Kriging model that they had developed by direct interpolation of DoC within each core. In addition to estimates of DoC at unsampled grid cells, this Kriging model also provided statistical estimates of uncertainty in the interpolated DoC values.

GE developed maps of DoC based on several percentiles of the Kriged uncertainty distribution and found relatively large differences between the percentiles of the Kriged distribution. For example, in the description of uncertainties in the Southern Portion of East Rogers Island (QEA, 2005; p 5-11 to 5-12), GE described an area where the 50$^{th}$ and 84$^{th}$ percentiles of the distribution differ by one foot, and the 50$^{th}$ and 95$^{th}$ percentile estimates differ by 2 feet (QEA, 2005; Figures 5-57 to 5-61). Similarly, in West Rogers Island, the 50$^{th}$ and 95$^{th}$ percentiles differed by approximately 2 feet as well.

Even with these identified uncertainties in interpolated values, GE elected to design to the 50$^{th}$ percentile without considering the problems that uncertainties of this magnitude would likely cause. By selection of a deterministic interpolation procedure that had similar uncertainties to that forecasted by Kriging[1], GE asserted that Kriging was an inappropriate technique, rather than draw what EPA believes to be a more accurate conclusion that the quality and quantity of DoC data were inadequate to constrain estimates of the DoC surface regardless of the interpolation method. Estimates of interpolation error based on the Kriging analysis indicated common errors of 1 foot or more, particularly in areas with deeper PCB deposits. The selected interpolation procedure contained no means to mitigate this uncertainty, and therefore GE proceeded as though measured and interpolated DoC were known.

Clay Layer Delineation

In some areas near Rogers Island, SSAP cores encountered the glacial Lake Albany clay (GLAC) layer that was anticipated to form a natural bottom to the vertical extent of PCB contamination. In discussions of how this might be incorporated into the design it was suggested that dredging contractors might be instructed to dredge until GLAC was encountered, rather than specifying a dredge elevation that was otherwise less certain. GE elected to hand contour the elevation of GLAC and instructed contractors to dredge to this approximated elevation. In practice, the uncertainty in this approach was high and GE's contractor dredged significant amounts of clay in efforts to attain the elevation estimate before the practice was corrected. In this situation, acknowledging the uncertainty in the elevation of the GLAC and

[1] Cross validation analysis conducted by EPA and GE indicated that uncertainties in DoC at unsampled locations were, on average, approximately 1 foot.

directing dredgers to dredge these areas until contact with GLAC was verified would have been more effective.[2]

Incorporation of the Navigation Channel Dimensions

In portions of the Phase 1 project, the dredging footprint overlapped the navigation channel that had been previously dredged to a water depth of approximately 14 feet, or approximately 103 feet in elevation. In these areas, the dredging design elevation based on analytical chemistry in cores and the 17-layer interpolation averaged 105 feet, approximately 2 feet shallower than the 14-foot depth and 3 feet shallower than the average final dredge elevation of 102 feet (Table 1). These 3 feet of additional dredging depth were achieved through multiple passes of 6 to 12 inches per pass, resulting in significant increase in time required to close CUs. Increasing the design elevation by 4 inches in these "deep" deposits would have no substantive effect on the rate of CU closure and is questioned by EPA. GE is not incorporating the high level of uncertainty in modeled DoC intended to characterize the design elevation in the navigation channel. Hopefully better coring data will partially correct for this.

This uncertainty could be greatly reduced by incorporating a conceptual model that assumes PCB contamination extends to the depth that any areas were dredged to, either just prior to or during the onset of PCB releases and have not been dredged to since then. It is recommended that GE design the dredge elevations based on 14 to 15 feet of water depth in areas known to have been dredged for navigational purposes. EPA recognizes that GE is reluctant to ignore analytical data in these areas; however, based on the Phase 1 experience, contamination extended to at least the 102-foot elevation and as deep as the 99-foot elevation (NAVD 88). It is now clear that cores failing to fully penetrate and recover sediments from these elevations were of little value in ascertaining the proper depth of cut. Given that these Phase 1 delineation errors were measured in feet, it is difficult to accept an argument that an additional 4 inches of overdredge tolerance will improve performance in these thicker deposits in Phase 2 dredging.

### 1.4. Diagnostic Testing Performed Prior to Dredging (2004-2005)

EPA performed diagnostic tests of GE's multiple layer DoC model called cross-validation. Cross-validation is a technique wherein individual Level 1A (complete) cores were removed from the analysis and the model was used to predict DoC at the dropped location where DoC had been measured. This analysis was also conducted on a Kriging model that was similar to that described by GE in the Phase 1 DAD. The results of this analysis are shown in Figure 1. A primary objective of the analysis was to evaluate model performance separately for SSAP locations with measured DoC less than or equal to 24 inches in comparison with those locations with measured DoC greater than 24 inches.

Figure 1 shows side by side box plots of the cross-validation errors. The lower and upper limits of the boxes represent the 25$^{th}$ and 75$^{th}$ percentiles of the error distribution. The red horizontal lines represent the 50$^{th}$ percentile (median) error. Negative errors represent under-prediction and positive errors represent over-prediction of DoC by the interpolation models. Figure 1 indicates that for thicker deposits, DoC was understated by 6 inches or more in 75 percent of locations and understated by 15 inches or more at 25

[2] In addition to the problems with the conceptual approach, the designers seem also to have interpreted any reference to clay in the sediment classifications as evidence of GLAC. In fact, such a reference may have simply represented a judgment call by the field classifier of the fraction of clay-sized particles in a sample, rather than evidence of a geological horizon.

percent of locations. In contrast, in shallow deposits, DoC was overstated (*i.e.*, predicted elevation was deeper than measured elevation) in 75 percent of locations. This indicates that there was substantial uncertainty in interpolated DoC that if left uncompensated would be expected to cause frequent failure of the Residuals Performance Standard. Roughly speaking, failure to adequately identify DoC in 25 percent of locations could be expected to cause an approximate 25 percent failure rate by area. Importantly, due to the nature of preferential representation of shallower deposits, actual performance in deeper deposits was anticipated to be worse than estimated.[3] With this understanding, underestimation of DoC in thicker deposits is clearly of greater magnitude and frequency than that in thinner deposits — the magnitude of errors being even greater than estimated from Level 1A cores.

GE's proposal to add 4 inches to DoC in thicker deposits and 6 inches in thinner deposits runs contrary to results forecasted in 2005 and subsequently verified by Phase 1 dredging in 2009. Uncertainties are clearly greatest in thicker PCB deposits and, therefore, EPA disagrees with the GE proposal to hedge the DoC more in shallow deposits than in thick deposits. It is likely that this approach will result in removal of more non-target sediment relative to potential increase in the percentage of target sediment removed.

### 1.5. Phase 1 Dredging Performance (2009)

Actual performance of the design in Phase 1 dredging resulted in large deposits of unexpected PCB mass and volume below design elevations which, according to the Peer Review Panel, were due in large part to improper characterization of the PCB inventory (*i.e.*, nature and extent of PCB contamination). The Peer Review Report (Bridges, *et al.*, 2010) stated:

> *"The Residuals EPS was not achieved in Phase 1. Residuals management required multiple production passes (not anticipated in the EPS) and the CUs were open longer than intended. The Residuals EPS was not truly tested as envisioned in Phase 1, mainly because inventory was improperly characterized and the EPS assumed that all inventory would be removed with a maximum of 2 passes, followed by additional passes to remove dredge-generated residuals.*
>
> *The incomplete characterization of inventory was attributed primarily to problems with the delineation of the DoC in much of the river, which was rooted in problems with sediment core data, including lack of absolute vertical control on the DoC, poor core recoveries, and inability to characterize the entire soft sediment column by coring to till. Consequently, core sample results fed into the Terrain Model provided inadequate representation of the DoC, and dredging to the Terrain Model DoC fell short in all CUs."*

During implementation of Phase 1 dredging, deviations between the final dredge depths and the design DoC averaged 11 inches across all CUs, and were positively associated with design DoC. Figure 2 shows that in Level 1A cores, the thickness of additional dredging increased with the estimated design DoC. It can also be seen that the number of Level 1A cores declines with increasing design depth because the likelihood of penetrating and retaining sediments that capture the DoC is a decreasing

[3] Level 1A cores preferentially represent the thinner deposits and areas with more cohesive sediments where cores are more likely to penetrate and retain the full contaminated sediment layer. Because of this bias, the magnitude to which modeled DoC is underestimated in thicker deposits is itself understated by this analysis of Type-1A cores.

function of PCB deposit thickness. As described previously, this is causes estimates of errors in thicker deposits to be understated. Actual performance is expected to be worse than the estimate.

Level 1A cores represent a biased sample of sediments that over-represent thinner deposits and soft cohesive sediments, each of which is more likely to produce conditions conducive to penetration and recovery of sediments. Accurate determination of DoC is more difficult in thicker deposits, because the probability of retaining a complete sediment profile declines with the thickness of sediment, resulting in fewer PCB measurements near the bottom of the contamination layer in thicker deposits. The primary driver of uncertainty in estimated DoC elevations is sample density and this reduction in sample density precisely in the vicinity of the EoC causes increased uncertainty in these areas. Therefore DoC uncertainty is necessarily greatest in areas where soft sediments are thick. GE's proposal to add less overdredge tolerance in thicker deposits runs counter to this controlling physical mechanism that manifested itself in severe understatement of DoC in Phase 1 dredging areas with thicker deposits.

### 1.6. Coring Results for Phase 2 Areas (2010)

A sampling program was implemented in 2010 with the objective to collect sediment cores that penetrate more deeply and recover higher percentages of sediment than was achieved by the SSAP program. The new coring program is termed the Supplemental Engineering Data Collection (SEDC) program, and coring locations are focused primarily on areas with incomplete cores with a subset of locations near existing complete cores. Figure 3 provides a plot of Level 1A SEDC cores vs. neighboring Level 1A cores from the SSAP program that are within 20 feet of the SEDC core. The plot also includes a regression line with a slope of 0.88, indicating that the SEDC DoC was, on average, 12 percent less than DoC in neighboring SSAP cores.

In addition to this moderate bias between the SEDC and SSAP programs there was also substantial random variation between SEDC and SSAP measures. For example for SSAP cores with a DoC of 30 inches, the SEDC DoC measurements ranged from 12 to 42 inches - an approximately 12-inch range of values. The standard deviation of these differences was 8.2 inches, which could be considered to be an estimate of the nugget effect of the DoC variogram. These results are consistent with previous variogram and cross validation analyses that indicated that nugget effect was approximately 9 to 12 inches and that cross-validation, root-mean-squared errors were approximately 10 inches.

In Figure 3, it is also noticeable that the frequency of Level 1A SSAP cores which can be paired with a SEDC core decreases dramatically with deposit thickness. This again reflects the biasing effect of sediment thickness on the likelihood of penetrating and recovering a complete sediment profile in thicker deposits.

The primary conclusion that can be drawn from these data is that substantial uncertainty exists in essentially co-located cores. Uncertainty in DoC at unsampled locations is expected to be generally higher than observed in these closely spaced pairs. While it is anticipated that an increased percentage of complete cores will improve the delineation in Phase 2 areas, the Peer Reviewers' expectation that a highly accurate DoC could be achieved through replacement of incomplete cores with complete cores is not supported by these data. This suggests that dredge design elevations should be developed with an adequate level of compensation for uncertainty to provide a reasonable expectation that the 1-mg/kg total PCB elevation is frequently achieved in the first dredging pass. Given the magnitude of errors observed in

SEDC-SSAP comparisons it is unlikely that deepening the dredging elevations by 4 to 6 inches will substantively improve dredging performance, particularly in thicker deposits where GE proposes to target the DoC less aggressively.

### 1.7. Summary and Conclusions

Additional dredging in the Phase 1 CUs was necessary because the design cut lines understated the true DoC. The relevance and consequences of uncertainty in the DoC measurements was a point of disagreement between EPA and GE over the course of the remedial design with EPA warning that "underestimating DoC would lead to additional re-dredging to remove inventory." Phase 1 results validate EPA's concern regarding DoC, as the average thickness of additional dredging required averaged greater than 1 foot and ranged to more than 9 feet. Underestimation of the DoC resulted in significant re-dredging to remove inventory and not residuals. This also resulted in multiple re-dredging passes which likely adversely impacted resuspension. Since it is clear that GE considered its design sufficiently robust to deal with DoC uncertainty during implementation, it is not appropriate to consider the consequences of this design flaw as an unexpected impediment to productivity, nor to attribute the need to re-dredge a flaw of the Residuals Standard.

Based on consideration of four lines of evidence described above, EPA concludes that GE's proposal to maintain the same methods for interpolation of DoC with the addition of 4 and 6 inches to the DoC in thick and thin deposits, respectively, is unlikely to achieve the "high-precision characterization of the DoC elevation" recommended by the Peer Review Panel. Rather, based on experience with Phase 1 dredging, analysis of precision of the GE DoC model, and analysis of new SEDC core data, EPA must conclude that an overdredge tolerance of a relatively small amount in thicker deposits, and potentially too much in thin deposits where uncertainties are smaller, could result in re-dredging and capping rates similar to those observed in Phase 1. This overdredge tolerance approach may also result in removal of unnecessary amounts of non-target sediment in thinner PCB deposits, ultimately with GE accruing costs both for removal of excess non-target sediment in places where DoC is reliably known and in re-dredging and capping costs in thicker deposits. With this realization, EPA is adopting a performance-based approach to design approval and implementation, with the expectation that adaptive management strategies will be required early within the 2011 dredging season.

In efforts to persuade GE to mitigate at least some of the uncertainties expected to compromise dredging performance in Phase 2 – Year 1 (2011), the following section describes a modified approach to development of the elevation of contamination for GE's consideration. The approach is not prescriptive, but does provide adequate detail to convey the underlying conceptual model that EPA believes is pervasive within the Phase 2 areas.

## 2. RECOMMENDATIONS FOR DEVELOPMENT OF ELEVATION OF CONTAMINATION SURFACES AND DREDGE PRISMS

Dredge prism development consists of two primary activities: development of surfaces corresponding to the elevation of contamination (which is the depth of contamination subtracted from the bathymetric surface), followed by incorporation of the physical characteristics of the river such as presence of debris, presence of structures like bridges, and other engineering considerations. Once developed these EoC surfaces can be "optimized" by subjecting them to statistical evaluation estimating the rate of successfully meeting performance criteria. Common performance criteria may include anticipated re-dredging rates, capping percentage, the volume of target and non-target sediments removed as well as the volume of target material left behind. This section describes how contamination elevation can be defined and how the development of specific prisms that incorporate this elevation can be developed and evaluated against performance criteria.

### 2.1. Method Overview

1. Identify conceptual model of physical process;
2. Determine accurate Elevation of Contamination (EoC) at core locations;
3. Calculate Depth of Contamination (DoC) at core locations;
4. Predict DoC at unsampled locations based on physical features and interpolation procedures;
5. Convert DoC predictions and uncertainty bounds to EoC by subtraction from bathymetry;
6. Use EoC and all available physical information to develop dredge prisms; and
7. Optimize DoC for performance criteria and consider modifications to final EoC.

### 2.2. Conceptual Model

Predicting a surface describing the EoC requires development of a conceptual model of deposition and erosion. Once defined, the appropriate processes (*e.g.*, thinning of deposits on steeper slopes) can be conceptualized and translated into design variables. The predicted EoC surfaces should be consistent with what is understood about sediment deposition. If adequate data are available to establish depositional and/or erosional relationships quantitatively, this approach can significantly improve the accuracy and precision of EoC predictions. The general physical factors guiding conceptual model development include:

1. Sediments tend to deposit on mild slopes in a uniform manner. Significant bathymetric changes typically result from scour conditions.[4] Dredge prisms for anticipated mechanical dredging should consist of horizontal – or nearly horizontal – cut lines and stair-steps cuts where necessary, with dredge elevations that parallel the existing long flow bathymetry combined with stable side slopes perpendicular to flow. The predicted EoC should reflect a synthesis of all available information

[4] Sediments were more mixed in Phase 1 CUs around Rogers Island because of the influence of mud flows resulting from historical dam removal.

and should be constrained to be relatively horizontal to maintain consistency with the conceptual model.

2. Variability in EoC measurements represents a significant obstacle to accurately targeting the EoC at both sampled and unsampled locations. Because of the significant variability in EoC measurements, statistical procedures are necessary to quantify uncertainty at unsampled locations for incorporation in dredge prism design.

3. Areas which have been dredged previously (*e.g.*, the navigation channel) are expected to act as sediment accumulation areas, and in general it is expected that contaminated sediments will extend to previously-dredged elevations. In these areas, the EoC should be set at or below the historical dredged elevation. For example, in the navigation channel, the historically maintained water depth would have been 14 feet below mean low water as defined by the NYSCC (*i.e.*, 12 feet of required navigation draft plus 2 feet of additional dredging to facilitate the maintenance of the required draft).

Each CU may present its own unique circumstances that cause EoC measurements to deviate from the conceptual and mathematical models. This inconsistency must be corrected through engineering judgment during final design.

### 2.3. Procedure for Establishing EoC

EoC accuracy can be improved significantly by using absolute elevations for sediment coring results, establishing accurate EoC determinations at sample locations, and incorporating physical principles to guide dredge prism design at unsampled locations. The following is an outline describing a possible procedure to improve accuracy and provide uncertainty bounds for targeting EoC. Many of the techniques described below have been used on various sites, in particular the Fox River.

The method includes three components:

1. Measures to improve determination of EoC accuracy at sampled locations;

2. Procedures to incorporate physical principles into estimation of EoC at unsampled locations (*i.e.*, to interpolate core data); and

3. Techniques to integrate predicted EoC and all available physical information to establish final dredge prism design.

This approach will result in more accurate targeting of EoC, with associated uncertainty bounds. The results are intended to improve the accuracy of the design in order to support the one-pass approach recommended by the Peer Review Panel.

The purpose of the statistical analysis described herein is to produce a mathematical model that systematically synthesizes the physical processes into dredge prism design as much as possible, while also providing uncertainty bounds on estimated EoC. Integration uncertainty in the relationship between physical processes and EoC requires a mathematical relationship. This model, which can be displayed and manipulated in geographic information systems or computer aided design software, provides layers of information that can be sliced into cross sections and used in the design process. These layers will include estimated EoC, as well as upper and lower uncertainty bounds throughout each dredging certification unit (CU). These layers are incorporated in the design process to develop an appropriate EoC at sampled and unsampled locations. In areas where the model provides accurate predictions, a final EoC surface should only require adjustments in areas of obvious inconsistencies with the physical setting, *e.g.*, around bridge abutments.

### 2.3.1. General Approach for Establishing EoC at Core Locations

All core analytical results, including those for SSAP cores, should undergo a review to determine appropriate bounds for estimating EoC at each sampled location. Physical information, including sediment core recovery rates, proximity to anthropogenic features, elevation and slope of bathymetric surfaces and sediment type based on detailed examination of coring logs, as well as EoC at neighboring locations both along and across flow, should be considered. The combination of analytical and physical information should lead to identification of a “best” EoC for each core location.

### 2.3.2. Prediction of EoC

Upon development of an accurate EoC for each core location, these individual EoC values will be used to develop a mathematical model of the EoC surface that is consistent with depositional processes captured in the conceptual model and the analytical data. It is expected that all available analytical data of sufficient quality will be incorporated into the development of the mathematical model. The purpose of this mathematical model, as opposed to simpler qualitative alternatives, is to provide a basis from which uncertainty bounds can be developed for EoC at unsampled locations, as well as to provide a systematic way to synthesize analytical and physical data. The model for EoC will be based on mathematical relationship between DoC (obtained by subtraction of EoC from the mud line elevation as determined from the best available bathymetry data closest to the date of core collection) and physical variables, including the geometry of the river channel, slope of the bathymetric surface, proximity to anthropogenic disturbances, water velocity, shear stress and other data that are available at greater spatial density than the core data themselves.

The synthesized estimates should be based on a combination of regression and Kriging of regression deviations (*i.e.*, residuals), sometimes called Kriging with spatially varying mean. Alternative interpolations to Kriging can be considered provided that interpolated surfaces are consistent with the conceptual model governing sediment deposition and that quantitative prediction limits are available at unsampled locations.

### 2.3.3. Performance Optimization

Because EoC is known at a small fraction of the surface that is characterized (*i.e.*, one 4 inch diameter core per 80 foot spacing, the degree to which fluctuations in EoC are constrained by sample data determines the likely outcomes of design performance metrics. The statistical model used to develop the EoC surfaces can be used to generate synthetic surfaces that are consistent with the statistical properties of sample data and other known constraints such as EoC confining surfaces. These synthetic “maps” can be used to propagate uncertainty in the EoC surface into forecasts of performance metrics. This approach has been used successfully at other superfund sites to improve the cost effectiveness of remedial designs.

In the months following release of the Peer Review report, EPA met with GE on multiple occasions to discuss uncertainty in DoC that adversely impacted performance of the Phase 1 dredging effort. EPA reiterated their 2005 recommendation to develop a statistically based model, adequately rigorous, to quantify the distributions of important dredging performance metrics. Over this period, EPA also provided GE with a range of alternatives that could be used to quantify the likely effect of uncertainty in DoC and to a limited extent provide indications of the expected performance of the Phase 2 design. These suggestions ranged from back-of-the-envelope calculations to the more rigorous method that was implemented at the Fox River Superfund Site and is described below. GE's response to the suggestion of this rigorous statistical approach has been that an apparent absence of spatial correlation prohibits such an analysis at the Hudson River. EPA does not find this argument to be compelling because, as discussed further below:

1. the absence of spatial correlation in DoC would not preclude application of the recommended statistical procedures to evaluate expected dredge performance, and

2. GE previously demonstrated through extensive variogram analysis of the SSAP data (QEA, 2005) that DoC is spatially correlated within Thompson Island Pool, including East Griffin Island.

EPA understands that spatial correlation in the DoC distribution is weak; however, this is common of sediment data and does not preclude or diminish the importance of the analysis recommended by EPA. This weak correlation with high nugget effect is also characteristic of DoC measurements at the Fox River; yet analysis procedures used there, and recommended for the Hudson River, rigorously account for this common situation. Using these uncertainty models, forecast dredging rates at the Fox River were subsequently validated in actual dredging performance. Contrary to GE's arguments, the nature of the spatial correlation at the Hudson River is not an obstacle to conducting the recommended analysis but rather a precursor to results that may include high uncertainty in performance levels - weaker spatial correlation implies lower confidence in DoC at unsampled locations and correspondingly higher uncertainty in expected dredging performance. It is precisely the effect of these uncertainties that EPA is interested in understanding as they are likely to strongly influence the expected performance of GEs dredging design.

GE's analysis of spatial correlation based on data solely from the dredge areas at East Griffin Island, purportedly to "prove" that spatial correlation is absent, were flawed because this dataset is much too restricted to represent the results of sediment transport and deposition dynamics. The analysis also relied on a variogram that ignored the direction of flow in the river, a fundamental physical constraint that affects sediment transport and, thus, contaminant fate and correlation properties of the collected data. Constraining use of the data in this way predisposed the answer to understate the actual spatial correlation present - the analysis lacked adequate sample size to estimate the variogram parameters adequately. With over 3000 sampling locations in the Thompson Island Pool, it is inappropriate to derive this broad conclusion from one small subset of the data.

Previous variogram analyses conducted by GE and EPA in the East Griffin Island area, as well as elsewhere in the Thompson Island Pool, using the same SSAP data showed moderate levels of spatial correlation in the DoC data, with large variation over short distances (*i.e.*, nugget effect). It is not clear why GE's analysis of the SSAP data should have changed so dramatically over this period of time so as now to yield contradictory results.

It is recommended that such a procedure be implemented to evaluate and optimize the EoC surface developed using the preceding process. Details of this approach follow in the next section.

### 2.4. Statistical Design Optimization

The approach described above based on a combination of professional engineering and statistical modeling will result in a preliminary design surface that can be expected to properly capture a proportion of the PCB containing inventory while also removing some non-target inventory and leaving some target inventory behind. An optimized design surface would minimize non-target sediments removed and target-sediments left behind while balancing the overall cost of implementation. This is similar to the problem of optimizing the amount of a precious metal that can be recovered from *in-situ* ore deposits. It is well understood in the mining industry that statistical procedures are necessary to develop an optimized mine plan that maximizes recovered resources while minimizing removal of waste rock. Recently, similar procedures have been applied at other Superfund Sites (Kern *et al.*, 2009) to evaluate the relative benefits between increased resolution through sampling and dredging to a more conservative neat-line to improve the level of protectiveness of the remedial design. Figure 4 summarizes the simulation procedure used at the Fox River to evaluate alternative infill sampling programs and methods for hedging uncertainty in the designs. Patmont and Fox (2009) identified key adaptive management elements resulting from these evaluations including: 1) reduced removal of non-target sediment, and 2) fewer expensive re-dredging cleanup passes. They projected a net savings of over $6 million in the first year of the 10-year dredging project.

The simulation study provided designers with quantitative forecasts comparing performance of candidate neat-lines. Figure 5 shows a comparison of the amount of target and non-target sediment expected to be removed by two candidate neat-line development procedures. It can be seen that the neat-line based on indicator Kriging was forecast to remove significantly less non-target sediment than the preliminary 30-percent design prisms. The simulation studies also provided a means to forecast expected re-dredging rates which were estimated to be approximately 15 percent in OU3. Actual dredging in 2009 and 2010 resulted in approximately 17 percent re-dredging, well within the error bounds of the 15 percent estimate.

It is EPA's recommendation that GE's dredge elevations for the Hudson River should also be subjected to similar analysis for potential benefits of uncertainty-hedging strategies including strategic use of overdredge tolerance in high-uncertainty areas as well as possible increased sampling density to maximize the proportion of target sediments removed, while balancing the amount of non-target sediments removed and capping performance criteria. It is also EPA's belief that the efficiency of the current overdredge tolerance strategy proposed by GE could be improved through application of a similar rigorous analysis of uncertainties in the estimated EoC surfaces.

## 3. REFERENCES


Benaman, J. Connolly, J. and A. Clough, 2010. *Understanding Remedial Design Effectiveness and Uncertainty By Analyzing Design Data Quality*. Society of Environmental Toxicology and Chemistry. North America 31st Annual Meeting: Bridging Science with Communities, Oregon Convention Center | Portland, Oregon. 7–11 November 2010.

Bridges, T., R. Fox, P. Fuglevand, G. Hartman, V. Magar, P. Schroeder, T. Thompson, and SRA International, Inc. 2010. *Hudson River PCBs Site, Peer Review of Phase 1 Dredging, Final Report.* September 10, 2010.

General Electric Company. 2010. *Phase 1 Evaluation Report, Hudson River PCBs Superfund Site*. Prepared for General Electric Company Albany, New York by Anchor QEA, LLC, Glens Falls, New York, and ARCADIS East Syracuse, New York. March 2010.

General Electric Company. 2005. *Phase 1 Intermediate Design Report, Hudson River PCBs Superfund Site.*

Kern, J.W., J.R. Wolfe, and N. Barabas. 2009. *Sampling Density for Refinement of Dredge Prism Design in the Lower Fox River, Wisconsin, OU 2-5*. Fifth International Conference on Remediation of Contaminated Sediments. February 3, 2009.

Patmont, C. and R. Fox. 2009. *Adaptive Management to Improve Sediment Cleanup on the Lower Fox River.* Fifth International Conference on Remediation of Contaminated Sediments, February 3, 2009.

Quantitative Environmental Analysis, LLC. 2005. *Hudson River PCBs Site, Phase 1 Dredge Area Delination Report*. February 28, 2005. Prepared on behalf of General Electric Company, Albany NY.

USEPA. 2002. *Record of Decision and Responsiveness Summary for Hudson River PCBs Site.* February, 2002.

| **Table 1. Design and Final Dredge Elevations at SSAP locations in Navigation Channel (CUs 1-8, 17 and 18)** | |
|---|---|
| **Summary** | **Value (Feet)** |
| Average Design Elevation | 105.1 |
| Average Final Elevation | 101.9 |
| Historic Channel elevation | 103.0 |
| | |
| Bias (Final-Design) Elevation | -3.3 |
| Bias (Final-102) Elevation | 0.0 |
| | |
| Standard Deviation of (Final-Design) Elevation | 2.8 |
| Standard Deviation of (Final-102) | 3.3 |
| | |
| Note: Final elevation was deeper than 103 feet at 75% of SSAP core locations. | |

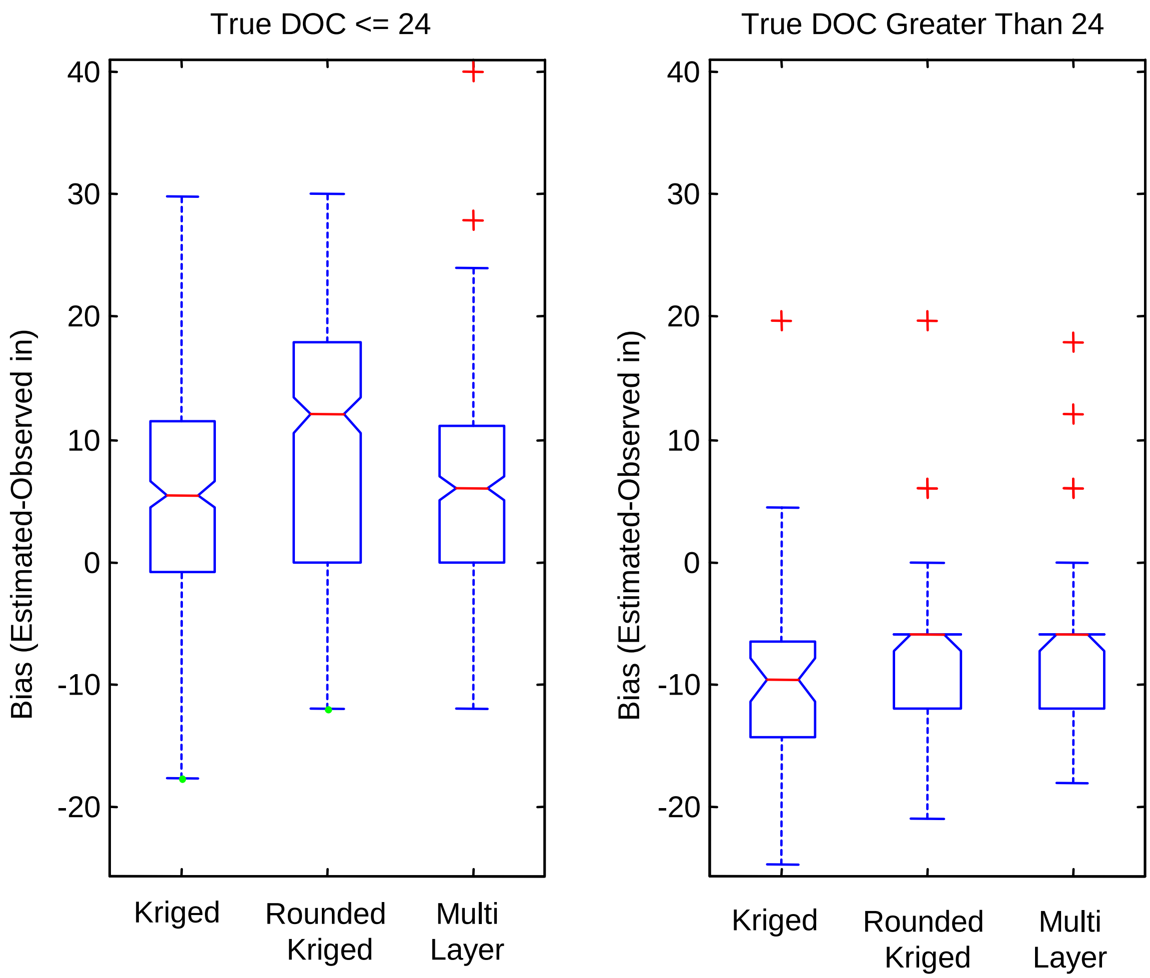


Figure 1. Estimated depth of contamination for sediments with DoC later proven to be less than or equal to 24 inches (Panel A) and sediments with DoC later proven to be greater than 24 inches (Panel B).

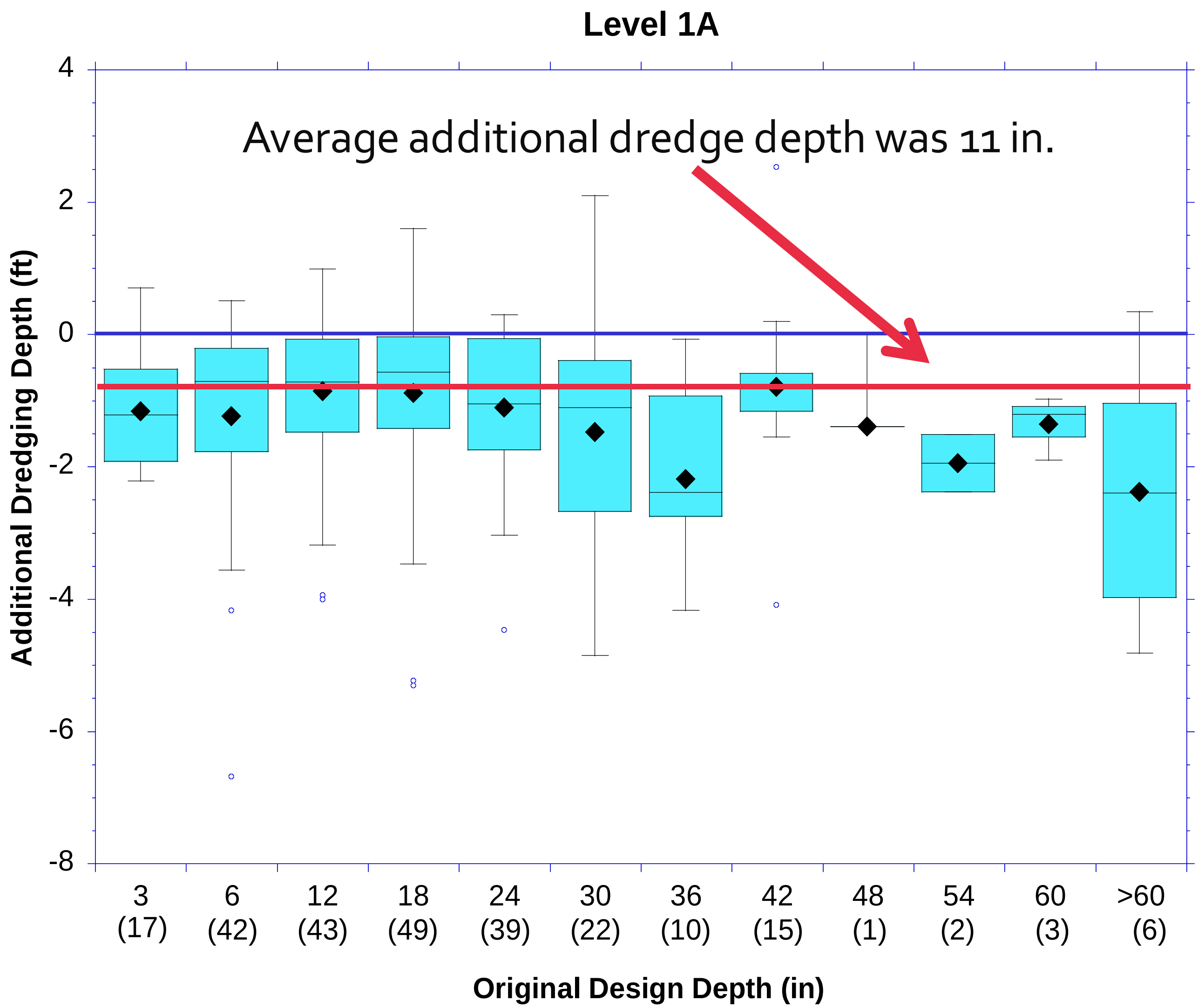


Figure 2. Distribution of additional dredge depth as a function of design dredge depth. Values in parentheses represent sample counts of complete (Level 1A) cores.

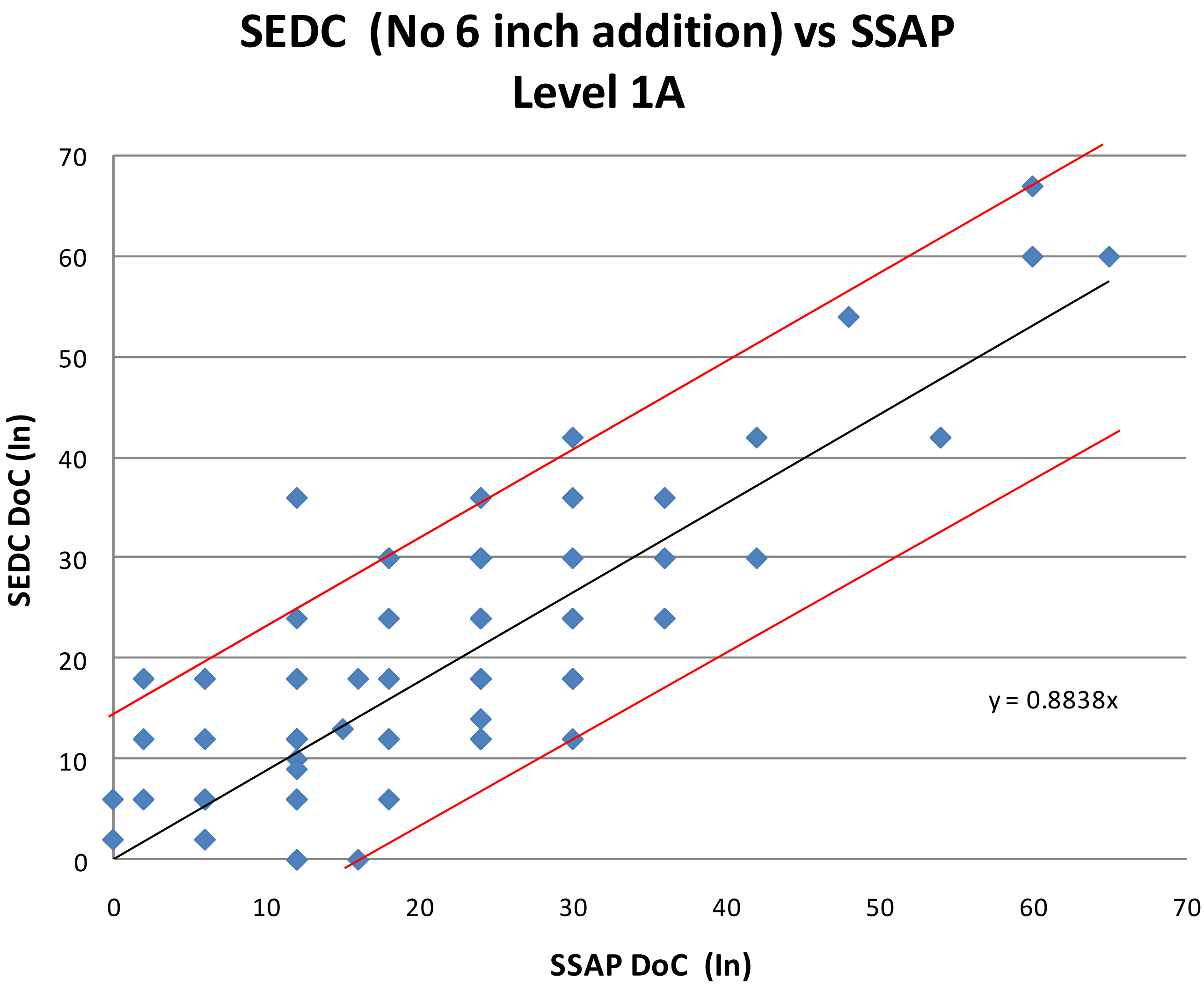


Figure 3. DoC in matched pairs of Level 1A SEDC and Level 1A SSAP cores within a 20 foot buffer.

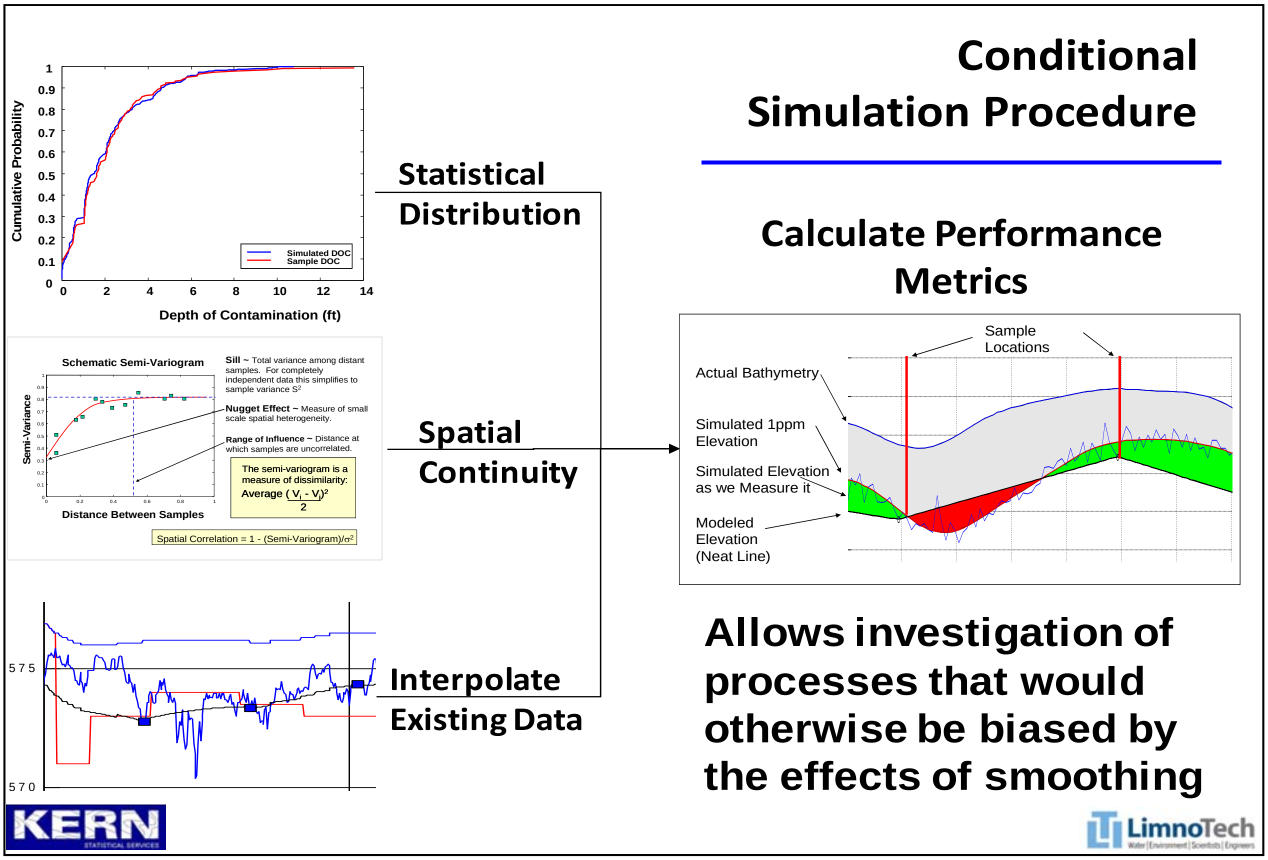


Figure 4. Simulation procedure for evaluating benefits of sample density and levels of conservatism in drawing dredge boundaries. (Figure excerpted from Kern, *et al.*, 2009).

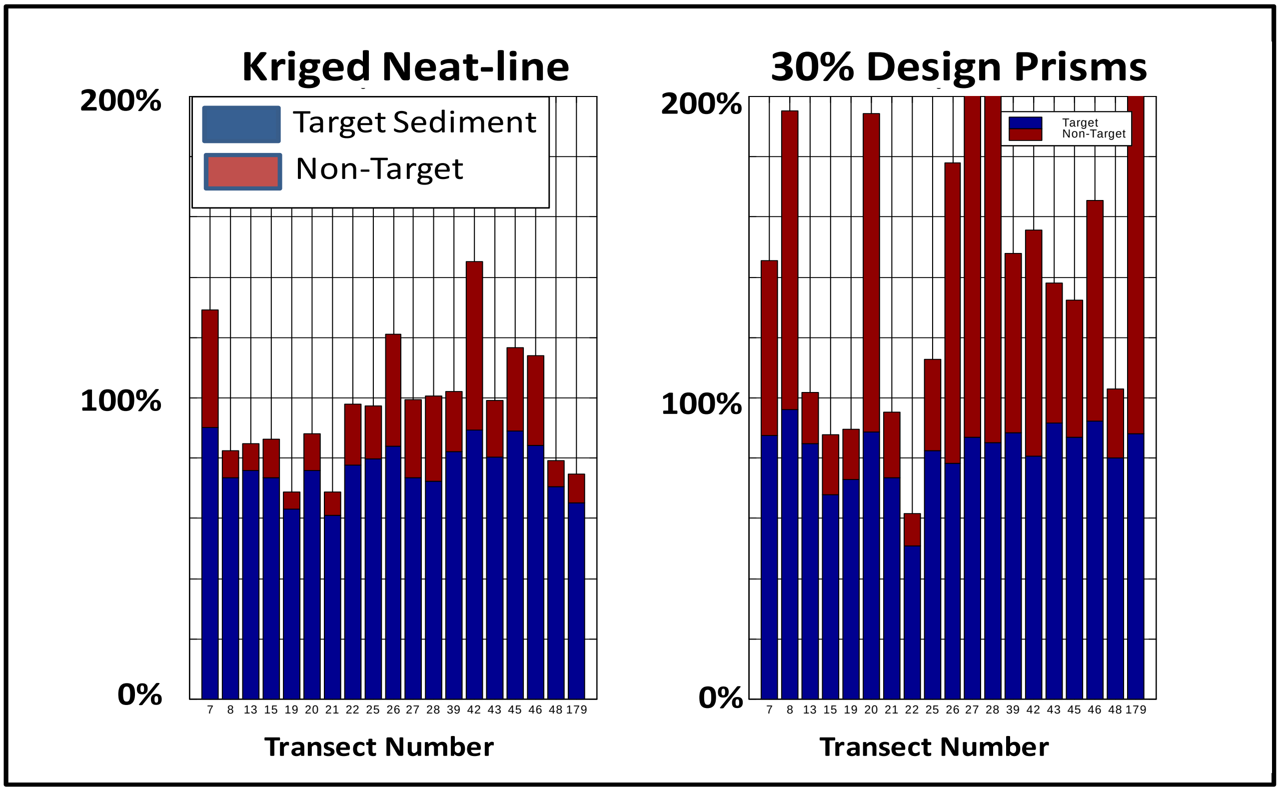


Figure 5. Comparison of Kriged neat-line and prism-based sediment dredging designs at the Fox River, WI. The Kriged neat-line was forecast to remove nearly the same volume of target sediment, while removing much less non-target sediment.